# On the Capacity of Oblivious Cloud Relay Access Networks

Shlomo Shamai (Shitz)

Technion Institute of Technology, Israel

Joint work with: Inaki Estella Aguerri (Paris Research Center, Huawei France)
Abdellatif Zaidi (Université Paris-Est, Paris) and Giuseppe Caire (TUB, Berlin)

sshlomo@ee.technion.ac.il

King's College, London

3 April, 2018

Supported by the European Union's Horizon 2020 Research And Innovation Programme: no. 694630.

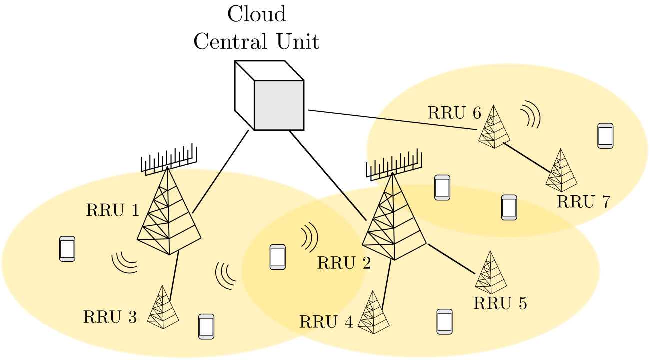
# Outline

1. **Introduction**
2. **Uplink Cloud RAN with Oblivious Processing**
   - Oblivious Relaying Schemes
   - Capacity Region of a Class of DM Channels
   - Capacity Region of Gaussian MIMO Channels with Gaussian Inputs
3. **Unconstrained Uplink Cloud RAN**
   - Structured coding
   - On the Oblivious Processing Constraint
4. **Concluding Remarks**

# Outline

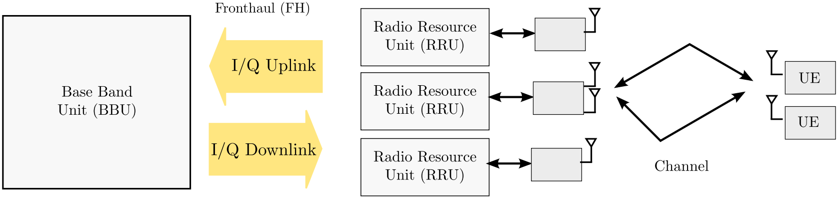
1. **Introduction**
2. **Uplink Cloud RAN with Oblivious Processing**
   - Oblivious Relaying Schemes
   - Capacity Region of a Class of DM Channels
   - Capacity Region of Gaussian MIMO Channels with Gaussian Inputs
3. **Unconstrained Uplink Cloud RAN**
   - Structured coding
   - On the Oblivious Processing Constraint
4. **Concluding Remarks**

# Introduction




- Cloud radio access network (C-RAN) architecture:
  - Heterogeneous dense networks;
  - Base stations (BSs), macro, pico, femto, operate as radio units (RUs);
  - Baseband processing takes place in the "cloud" or a central unit (CU).

# Distributed C-RAN Architecture




- Base Band Unit (BBU)
  - Encoding/decoding of the messages of User Equipment (UEs).
- Radio Rersource Unit (RRU)
  - Radio processing and analog-to-digital conversion.
- Fronthaul/Backhaul links (FH)
  - Carries complex inphase -quadrature (IQ) baseband signals.

# Advantages and challenges of the structure

**Advantages of the structure**

- Higher spectral efficiency: Effective interference mitigation via joint baseband processing (e.g. CoMP in LTE, 5G);
- Dense deployment with enhanced indoor coverage;
- Allows a cost effective way to deploy and upgrade wireless platforms;
- Flexible radio and computing resource allocation.

**Key challenge:**

- Large amount of I/Q data to be transferred over the FH between BBU and RRU *[Andrews et al JSAC14]*

# Cloud Radio Access Networks

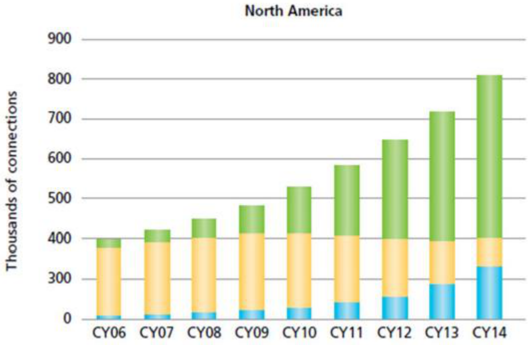



Figure: The distribution of backhaul connections for macro BSs (green: fiber, orange: copper, blue: air) *[Segel and Weldon]*.

- mmWave front/backhauling for 5G *[Ghosh 13] [Checko et al 15][Fujitsu 15]*
- Copper (LAN cable) for indoor coverage *[Lu et al 14]*

# Why Compression?

- Common public radio interface (CPRI) standard based on analog-to-digital (ADC)/digital-to-analog (DAC).

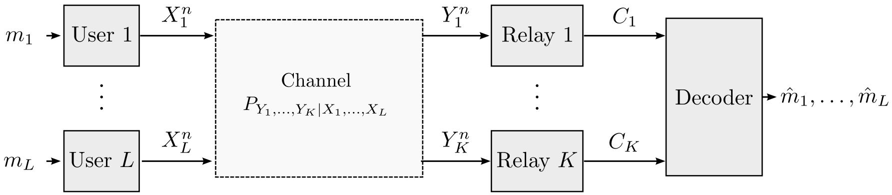
| Parameters | Settings |
|---|---|
| Sectors | 3 |
| LTE Carriers | 5 |
| Bandwidth | 100MHz |
| MIMO | $2 \times 2$ |
| Bits per IQ | 15 bits |
| Protocol | LTE-A |

**Required Fronthaul Throughput**

$$N_{\text{ant}} \times \text{samples/sec} \times \text{bits/sample} = 13.8\text{Gbps}$$

- Higher rate than supported by standard optical fibers ($10$Gbps).
- Need for compression on the fronthaul!
- "Death by Starvation ? : backhaul and 5G," *[Lundqvist, CTN-Sep. 2015]*

# Uplink Cloud RAN




- Multiple access relay channel in which $L$ users communicate with a common destination through $K$ relay nodes.

- Decoder interested in $\hat{m}_1,\ldots,\hat{m}_L$ such that, for $n$ large enough,

$$\Pr\{(m_1,\ldots,m_L) \neq (\hat{m}_1,\ldots,\hat{m}_L)\} \to 0$$

- The capacity region of this model is still to be found
  - problem open even in seemingly simpler cases, e.g., one user and two relays (the diamond channel), parallel Gaussian relay channel [Schein-Gallager '00]).

# Relay Operations

- Main difficulty is in characterizing the optimal relay operation:
  - **Decode-and-Forward:** *[Cover-ElGamal'97], [Kramer-Gastpar'05] ...*
  - **Compute-and-Forward:** *[Nazer-Gastpar'11], [Nazer et al'12], [Hong-Caire'13]...*
  - **Compress-and-Forward:** *[Sanderovich et al'09], [Park et al'13], [Zhou et at'13]...*
  - **Noisy Network Coding:** *[Lim et al'11]...*
  - **Others:** Amplify and Forward, Partial-Decode-Compress-and-Forward *[Cover-ElGamal'97]*, Compute-Quantize-and-Forward *[Estella-Zaidi'16]...*
- Relaying operations can be divided into:
  - Non-oblivious: relays aware of the users' codebooks (modulation, coding...) at all time, e.g., DF, CompF
  - Oblivious (or Nomadic): *[Sanderovich et al'08]* relays operate without knowledge of the users' codebooks, e.g., CoF, NNC
- Oblivious processing motivated mainly by practical constraints

# Outline

# Randomized Encoding as a Model for Obliviousness

- Formally, obliviousness of the relays to actual codebooks is modeled through randomized encoding *[Sanderovich et al'08], [Lapidoth-Narayan'98]*:
- Encoding function at transmitter

$$\phi^n : [1, |\mathcal{X}|^{n2^{nR}}] \times [1, 2^{nR}] \to \mathcal{X}^n$$

which maps a a codebook index $F \in [1, |\mathcal{X}|^{n2^{nR}}]$ and a message $M \in [1, 2^{nR}]$ into a codeword $X^n(F, M) = \phi^n(F, M)$.

- The pair $(p_F, \phi^n)$ must satisfy

$$\text{Prob}[X^n(F, M) = x^n] = \prod_{i=1}^{n} p_X(x_i)$$

for some $p_X(x)$, $x \in \mathcal{X}$, where Prob$[\cdot]$ is calculated with respect to

$$p_{F,M}(f, m) = p_F(f) \cdot 2^{-nR}$$

# Randomized Encoding as a Model for Obliviousness (Cont.)

- Oblivious nodes, e.g., relays, are not aware of the actual users' codebooks

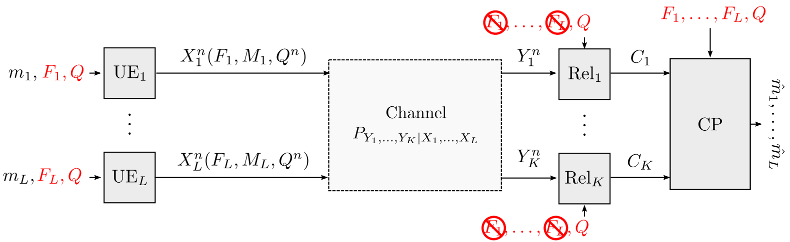
$$\phi_k^n : \mathcal{Y}^n \to [1, 2^{nC}]$$

- Non oblivious nodes, e.g., decoder, are aware of the actual users' codebooks

$$\psi^n : [1, |\mathcal{X}|^{n2^{nR}}] \times [1, 2^{nC}] \to [1, 2^{nR}]$$

- Although multi-letter expressions are precluded, still general enough to encompass single-letter superposition coding and other schemes

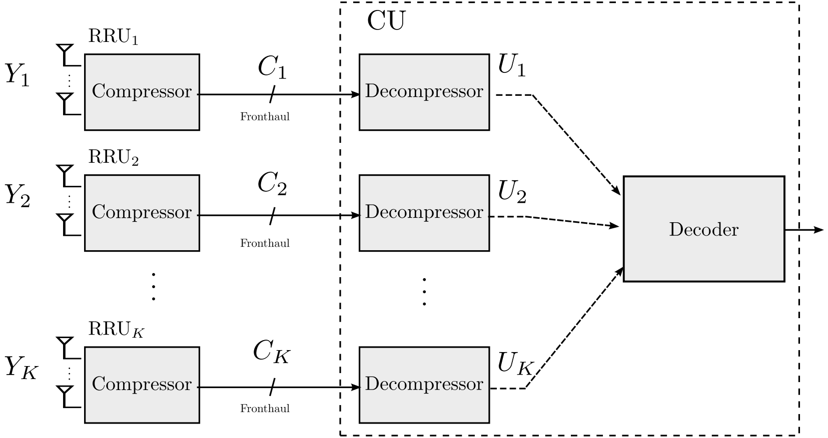
# Oblivious Relay Processing with Enabled Resource-sharing




- Resource-sharing random variable $Q^n$ available at all terminals *[Simeone et al'11]*
- $Q^n$ way easier to share, (e.g., on/off activity )
- Memoryless Channel: $P_{Y_1,\dots,Y_K|X_1,\dots,X_1}$
- User $l \in \{1,\dots,L\}$: $\phi_l^n : [1, |\mathcal{X}_l|^{n2^{nR_l}}] \times [1, 2^{nR_l}] \times \mathcal{Q}^n \to \mathcal{X}_l^n$
- Relay $k \in \{1,\dots,K\}$: $\phi_k^n : {\mathcal{Y}_k}^n \times \mathcal{Q}^n \to [1, 2^{nC_k}]$
- Decoder:

$$\psi^n : [1, |\mathcal{X}_1|^{n2^{nR_1}}] \times \cdots \times [1, 2^{nC_K}] \times \mathcal{Q}^n \to [1, 2^{nR_1}] \times \ldots \times [1, 2^{nR_L}]$$

# Point-to-Point Compression




**Point-to-Point Compression (PtP)**

- Received signals at different RRUs are separately compressed.

# Point-to-Point Compression (cont'd)

- A compression strategy exists if the fronthaul capacity $C_k$ satisfies

$$I(Y_k; U_k) \leq C_k,$$

- After recovering $(U_1, \ldots, U_K)$ jointly, the CP decodes the transmitted signals by the UEs $X_{\mathcal{L}}$, jointly. For all $\mathcal{T} \subseteq \{1, \ldots, L\}$

$$\sum_{t \in \mathcal{T}} R_t \leq I(X_{\mathcal{T}}; U_1, \ldots, U_K | X_{\mathcal{T}^c})$$

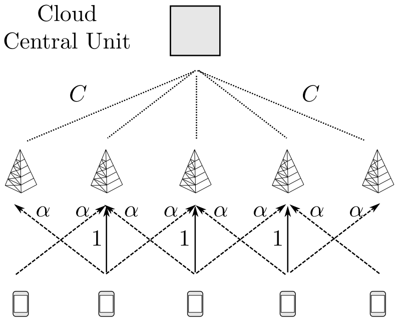
- A standard way of modeling compression at RRU $k$ is to adopt the Gaussian direct "test channel" *[El Gamal-Kim'11, Ch. 3]*

$$\mathbf{U}_k = \mathbf{Y}_k + \mathbf{Q}_k,$$

where $\mathbf{Q}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}_k)$ represents the quantization noise.

Advantage: Universality *[Lapidoth, IT'97]*

# Numerical example: Circular Wyner Model



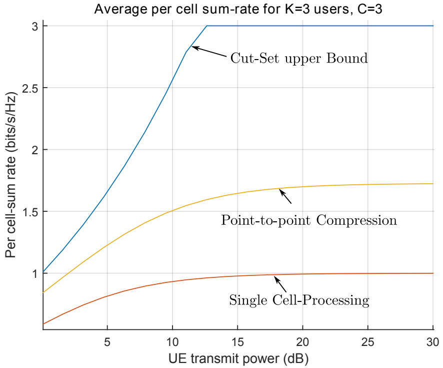

- Each cell contains a single-antenna and a single-antenna RU.
- Inter-cell interference takes place only between adjacent cells.

$$Y_k = \alpha X_{k-1} + X_k + \alpha X_{k+1} + N_k$$

where $N_k \sim \mathcal{CN}(0, 1)$

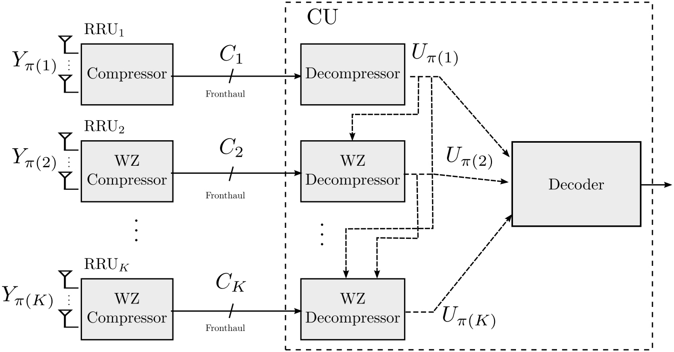
- All RRUs have a fronthaul capacity of $C$.

# Numerical example: Circular Wyner Model (cont'd)




- Single cell processing (Non-oblivious): Each cell decodes the signal of the in cell EU treating the other UEs as noise.
- Cut-set Bound *[Simeone et all' 12]*

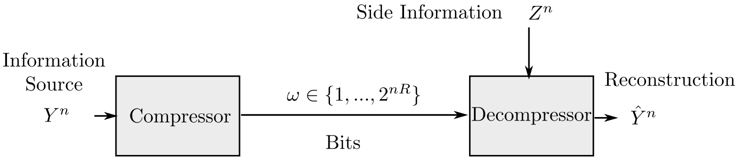
# Distributed Compression with Successive Wyner-Ziv




**Successive Wyner-Ziv (SWZ)**

- Account for correlation between received signals at different RRUs *[Sanderovich et al 09] [dCoso-Simoens 09] [Zhou-Yu 11]*
- $\pi : \mathcal{K} \rightarrow \mathcal{K}$ denotes a permutation of RRU's indices, $\mathcal{K} = \{1, \ldots, K\}$.

# Wyner-Ziv Source Coding Problem




- Wyner-Ziv Rate-distortion function:

$$R_{\mathrm{WZ}}(D) = \min_{P_{\hat{Y}|Y}} I(Y;\hat{Y}|Z) \quad \text{s.t. } \mathrm{E}[d(Y;g(\hat{Y},Z))] \leq D.$$

- Compression described by test channel $P_{\hat{Y}|Y}$. (Note that $\hat{Y} \multimap Y \multimap Z$).
- Gaussian source $Y \sim N(0,\sigma^2)$ and SI $Z$, with correlation coeff. $\rho$,

$$R_{\mathrm{WZ}}(D) = \frac{1}{2}\log^{+}\left(\frac{\sigma^2(1-\rho^2)}{D}\right).$$

# Distributed Compression with Successive WZ (cont'd)

- Using Wyner-Ziv Compression, a compressor exist if

$$I(Y_{\pi(k)}; U_{\pi(k)} | U_{\pi(1)}, \ldots, U_{\pi(k-1),Q}) \leq C_{\pi(k)}.$$

- After the quantized IQ signals $U_1, \ldots, U_K$ are recovered, the CU successively decodes signal $X_{\mathcal{L}}$ sent by all UEs in a particular order $\bar{\pi}$:

$$R_{\pi(l)} \leq I(X_{\bar{\pi}(l)}; U_1, \ldots, U_K | X_{\bar{\pi}(1)}^{\bar{\pi}(l-1)}, Q).$$

- Compression at RRU $k$ uses the Gaussian direct "test channel"

$$\mathbf{U}_k = \mathbf{Y}_k + \mathbf{Q}_k,$$

  where $\mathbf{Q}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}_k)$ represents the quantization noise.

- Universality aspects, *[Lapidoth-Narayan'98]*

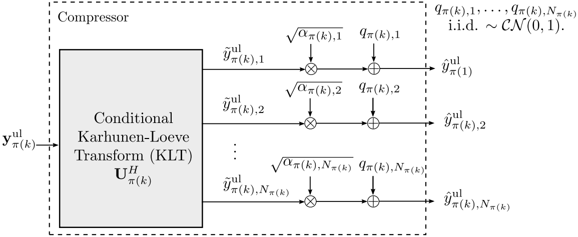
# Distributed Fronthaul Compression (cont'd)

- Sum-rate maximization problem with fronthaul capacity constraints is generally challenging.
- In *[Park et al TVT13]*, a block-coordinate optimization approach was proposed for successive WZ decompression case.
  - One optimizes the covariance matrices $\mathbf{\Omega}_{\pi(1)}, \ldots, \mathbf{\Omega}_{\pi(K)}$ following the same order $\pi$ employed for decompression.
  - At the $k$-th step, for fixed (already optimized) covariances $\mathbf{\Omega}_{\pi(1)}, \ldots, \mathbf{\Omega}_{\pi(k-1)}$, the covariance $\mathbf{\Omega}_{\pi(k)}$ is obtained by solving

$$\max_{\mathbf{\Omega}_{\pi(k)} \succeq 0} I(X_{\mathcal{L}}; U_{\pi(k)} | U_{\pi(1)}, \ldots, U_{\pi(k-1)})$$
$$\text{s.t. } I(Y_{\pi(k)}; U_{\pi(k)} | U_{\pi(1)}, \ldots, U_{\pi(k-1)}) \leq C_{\pi(k)}.$$

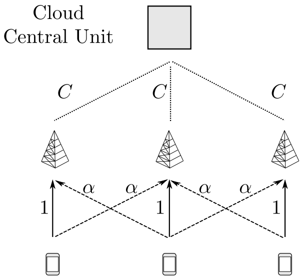
# Distributed Compression with Successive WZ (cont'd)

**Optimal Wyner-Ziv Compressor** *[dCoso-Simoens 09]*

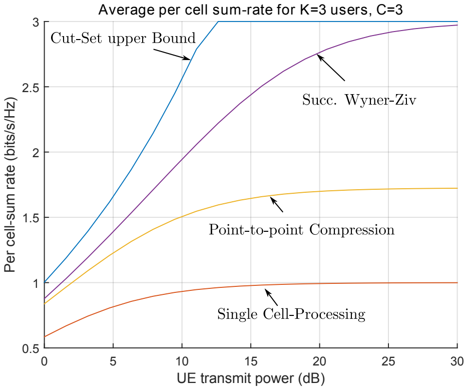



- Unitary transform $\mathbf{V}_{\pi(k)}^H$ decorrelates the received signal streams when conditioned on the side information signals $\mathbf{U}_{\pi(1)}, \ldots, \mathbf{U}_{\pi(k-1)}$
- Stream-wise multiplication by $\sqrt{\alpha_{\pi(k),1}}, \ldots, \sqrt{\alpha_{\pi(k),N_{r,\pi(k)}}}$ represents the compression rate allocation among the streams.
- Signals are compressed separately and quantization noises $\mathbf{Q}_{\pi(k),1}, \ldots, \mathbf{Q}_{\pi(k),N_{r,\pi(k)}}$ independent.

# Numerical example: Circular Wyner Model (cont'd)

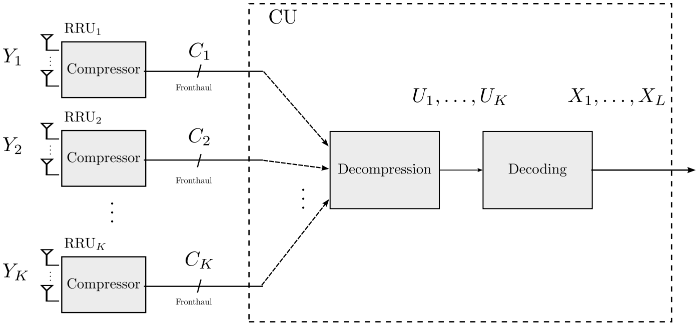



$$Y_k = \alpha X_{k-1} + X_k + \alpha X_{k+1} + N_k,$$
with $\mathrm{E}[X_k^2] \leq \mathrm{snr}$ and $N_k \sim \mathcal{CN}(0,1)$.




- The performance advantage of SWZ over PtP compression increases as the SNR grows larger.
- At high SNR, the correlation of the received signals at the RRU's is more pronounced.

# Separate Decompression and Decoding



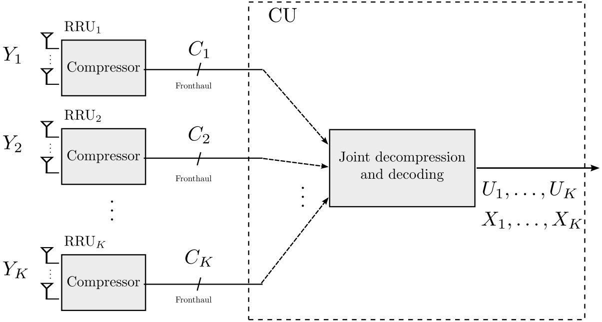

**Separate decompression and decoding (SDD)**

- Concatenation of Joint Decompression (Berger-Tung) and Joint Decoding

$$\sum_{t \in \mathcal{T}} R_t \leq I(X_{\mathcal{T}}; U_{\mathcal{K}} | X_{\mathcal{T}^c}, Q)$$

$$\sum_{s \in \mathcal{S}} C_s \geq I(Y_{\mathcal{S}}; U_{\mathcal{S}} | U_{\mathcal{S}^c}, Q)$$

# Joint Decompression Decoding




**Joint decompression and decoding (JDD)**

*[Sanderovich et al 09][Lim et al 11][Yassaee-Aref 11]*

- Potentially larger rates can be achieved with joint decompression and decoding (JDD) at the central unit *[Sanderovich et al 09]*.
- Now often seen as an instance of noisy network coding *[Lim et al 11]*.

# Joint Decompression Decoding (cont'd)

**Theorem**

*A rate tuple $(R_1, \ldots, R_L)$ is achievable if for all $\mathcal{T} \subseteq \mathcal{L}$, and all $\mathcal{S} \subseteq \mathcal{K}$,*

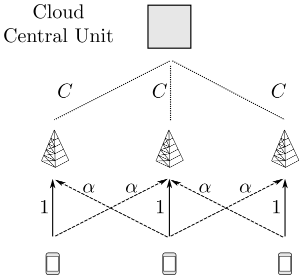
$$\sum_{t \in \mathcal{T}} R_t \leq \sum_{s \in \mathcal{S}} C_s - I(Y_S; U_S | X_{\mathcal{L}}, U_{S^c}, Q) + I(X_{\mathcal{T}}; U_{\mathcal{S}^c} | X_{\mathcal{T}^c}, Q),$$

*for some pmf*

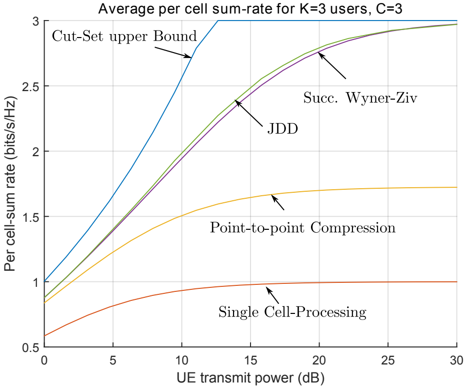
$$P_Q \prod_{l=1}^{L} P_{X_l|Q} \, P_{Y_1,\ldots,Y_K|X_1,\ldots,X_L} \prod_{k=1}^{K} P_{U_k|Y_k,Q}.$$

- Generalization of scheme from *[Sanderovich et al'09]* to $L$ users and channel $P_{Y_1,\ldots,Y_K|X_1,\ldots,X_L}$, with $Q$ the resource (time)-sharing parameter.
- Based on compress-and-forward à la Cover-El Gamal with joint decoding and decompression at the CP.

# Numerical example: Circular Wyner Model (cont'd)




$$Y_k = \alpha X_{k-1} + X_k + \alpha X_{k+1} + N_k,$$
with $\mathrm{E}[X_k^2] \leq \mathrm{snr}$ and $N_k \sim \mathcal{CN}(0,1)$.




- JDD is capacity achieving under oblivious processing.
- For this simple network, JDD does not provide much gain compared to SDD and SWZ.
  - Here, the schemes SDD and SWZ do not employ resource-sharing.

# On the Sum Rate of JDD

- Sum-rate $R_{\text{sum}} = \sum_{l=1}^{L} R_l$ achievable with JDD

$$R_{\text{JDD}}^{\text{sum}} = \min_{\mathcal{S} \subseteq \mathcal{K}} \left\{ \sum_{s \in \mathcal{S}} C_s - I(Y_{\mathcal{S}}; U_{\mathcal{S}} | X_{\mathcal{L}}, U_{\mathcal{S}^c}, Q) + I(X_{\mathcal{L}}; U_{\mathcal{S}^c} | Q) \right\},$$

  for some pmf $P_Q \prod_{l=1}^{L} P_{X_l|Q} P_{Y_{\mathcal{K}}|X_{\mathcal{L}}} \prod_{k=1}^{K} P_{U_k|Y_k,Q}$.

- Using properties of sub-modular functions, we show that SWZ achieves same sum-rate as JDD and SDD
- Note, however, that time-sharing is generally needed for the three to achieve optimal sum-rate!

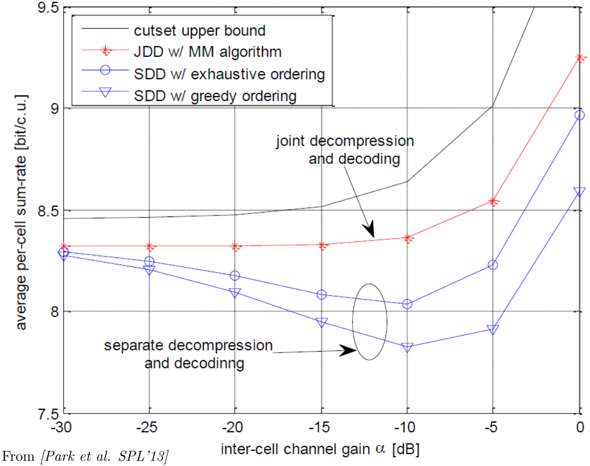
**Theorem**

*For any $P_{Y_1,\ldots,Y_K|X_1,\ldots,X_L}$, not necessarily satisfying $Y_k -\!\!\!\circ\!\!\!- X_{\mathcal{L}} -\!\!\!\circ\!\!\!- Y_{\mathcal{K}\setminus k}$ we have*

$$R_{\text{JDD}}^{\text{sum}} = R_{\text{SDD}}^{\text{sum}} = R_{\text{SWZ}}^{\text{sum}}$$

- In particular, for Gaussian MIMO channels recovers *[Zhou et al.'16]*
- In terms of rate region, JDD generally outperforms SDD.

# Numerical example: 3 Cell Uplink




From *[Park et al. SPL'13]*

- Optimizing over relay compression ordering $\pi$ improves performance.
- Resource-sharing required to achieve the sum-rate of JDD with SWZ.

# Main Capacity Results

Single-letter characterizations of:

1) *Capacity Region* of the Class of DM CRAN channels satisfying

$$Y_k^n -\!\!\circ\!\!- X_{\mathcal{L}}^n -\!\!\circ\!\!- Y_{\mathcal{K}\setminus k}^n,$$

2) *Capacity Region of Gaussian MIMO Channels with Gaussian Inputs*

- In particular, we show that Gaussian auxiliaries are optimal
- We study the role of Resource-sharing

3) *Inner and Outer Bounds* for General DM Model

# Capacity Region of a Class of CRAN Channels

**Theorem**

*For the class of discrete memoryless channels satisfying*

$$Y_k -\!\!\circ\!\!- X_{\mathcal{L}} -\!\!\circ\!\!- Y_{\mathcal{K}\setminus k}$$

*with oblivious relay processing and enabled resource-sharing, a rate tuple* $(R_1, \ldots, R_L)$ *is achievable if and only if for all* $\mathcal{T} \subseteq \mathcal{L}$ *and for all* $\mathcal{S} \subseteq \mathcal{K}$*,*

$$\sum_{t \in \mathcal{T}} R_t \leq \sum_{s \in \mathcal{S}} [C_s - I(Y_s; U_s | X_{\mathcal{L}}, Q)] + I(X_{\mathcal{T}}; U_{\mathcal{S}^c} | X_{\mathcal{T}^c}, Q),$$

*for some joint measure of the form*

$$P_Q \prod_{l=1}^{L} P_{X_l|Q} \prod_{k=1}^{K} P_{Y_k|X_{\mathcal{L}}} \prod_{k=1}^{K} P_{U_k|Y_k,Q}.$$

*with the cardinality of* $Q$ *bounded as* $|\mathcal{Q}| \leq K + 2$*.*

# Direct Part

Capacity region achievable with

- **Compress-and-Forward with Joint-Decompression-Decoding**
  - Generalization of scheme from *[Sanderovich et al'09]* to $L$ users
  - Based on compress-and-forward à la Cover-El Gamal with joint decoding and decompression at the CP.
  - Gaussian inputs are not optimal for finite capacity fronthauls
- Separate decompression-decoding not optimal in general
- **Noisy Network Coding**
  - Particular case of *[Theorem 1, Lim et al'11]*

Sum-rate achievable also with

- **Compress-and-Forward with Separate-Decompression-Decoding**
  - The CP decodes explicitly the compression indices first and then decodes the users' transmitted messages.

# Outline of Converse Part

- Define $\quad U_{i,k} \triangleq (J_k, Y_k^{i-1})$ and $\bar{Q}_i \triangleq (X_{\mathcal{L}}^{i-1}, X_{\mathcal{L},i+1}^n, \tilde{Q})$.

  Fano's Inequality: $\quad H(m_{\mathcal{T}}|J_{\mathcal{K}}, F_{\mathcal{L}}, \tilde{Q}) \leq \epsilon_n$ for $\mathcal{T} \subseteq \mathcal{L}$,
- **Upper bound on entropy term:** For $\mathcal{T} \subseteq \mathcal{L} = \{1, \ldots, L\}$,

$$H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, J_{\mathcal{K}}, Q^n) \leq \sum_{i=1}^{n} H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, \bar{Q}_i) - n\sum_{t\in\mathcal{T}} R_t \triangleq n\Gamma_{\mathcal{T}}$$

  - Follows from

$$\begin{aligned} n\sum_{t\in\mathcal{T}} R_t = H(m_{\mathcal{T}}) &= I(m_{\mathcal{T}}; J_{\mathcal{K}}, F_{\mathcal{L}}, \tilde{Q}) + H(m_{\mathcal{T}}|J_{\mathcal{K}}, F_{\mathcal{L}}, \tilde{Q}) \\ &\leq I(m_{\mathcal{T}}, F_{\mathcal{T}}; J_{\mathcal{K}}|F_{\mathcal{T}^c}, \tilde{Q}) + n\epsilon_n \\ &\leq H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, \tilde{Q}) - H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, J_{\mathcal{K}}, \tilde{Q}) + n\epsilon_n \end{aligned}$$

- Reminiscent of log-loss penalty criterion in multi-terminal source coding *[Courtade-Weissman'14]*:

$$H(X^n|J_{\mathcal{K}}) \leq \mathrm{E}[d_{\mathrm{LL}}(X^n; \hat{X}^n)] \simeq n(H(X) - I(X; \hat{X}))$$

# Outline of Converse Part (Cont.)

- **Bound on users' rates:** For $\mathcal{T} \subseteq \mathcal{L}$

$$n \sum_{t \in \mathcal{T}} R_t \leq I(m_{\mathcal{T}}, F_{\mathcal{T}}; J_{\mathcal{K}} | F_{\mathcal{T}^c}, \tilde{Q}) + n\epsilon_n \leq \sum_{i=1}^{n} I(X_{\mathcal{T},i}; U_{\mathcal{K},i} | X_{\mathcal{T}^c,i}, \bar{Q}_i) + n\epsilon_n$$

- **Bound on relays' rates:** For $\mathcal{S} \subseteq \mathcal{K} = \{1, \dots, K\}$

$$\begin{aligned} n \sum_{k \in \mathcal{S}} C_k &\geq \sum_{k \in \mathcal{S}} H(J_k) \geq I(X_{\mathcal{T}}^n, Y_{\mathcal{S}}^n; J_{\mathcal{S}} | X_{\mathcal{T}^c}^n, J_{\mathcal{S}^c}, \tilde{Q}) \\ &\geq \sum_{i=1}^{n} H(X_{\mathcal{T},i} | X_{\mathcal{T}^c,i}, U_{\mathcal{S}^c,i}, \bar{Q}_i) - n\Gamma_{\mathcal{T}} + I(Y_{\mathcal{S}}^n; J_{\mathcal{S}} | X_{\mathcal{L}}^n, J_{\mathcal{S}^c}, \tilde{Q}) \\ &= nR_{\mathcal{T}} - \sum_{i=1}^{n} I(X_{\mathcal{T},i}; U_{\mathcal{S}^c,i} | X_{\mathcal{T}^c,i}, \bar{Q}_i) + \sum_{k \in \mathcal{S}} \sum_{i=1}^{n} I(Y_{k,i}; U_{k,i} | X_{\mathcal{L},i}, \bar{Q}_i) \end{aligned}$$

where we used the upper bound on the entropy and the Markov chain

$$Y_k -\!\!\circ\!\!- X_{\mathcal{L}} -\!\!\circ\!\!- Y_{\mathcal{K} \setminus k}$$

# Memoryless MIMO Gaussian Model

- The channel output at relay node $k$ with $M_k$ antennas:

$$\mathbf{Y}_k = \mathbf{H}_{k,\mathcal{L}}\ [\mathbf{X}_1^T, \ldots, \mathbf{X}_L^T]^T + \mathbf{N}_k,$$

  where

  - User $l$ with $N_l$ antennas transmits $\mathbf{X}_l$ with $\mathrm{E}[\|\mathbf{X}_l\|^2] \preceq \mathbf{K}_l$
  - Relay $k$ with $M_k$ antennas
  - $\mathbf{H}_{k,\mathcal{L}} = [\mathbf{H}_{k,1}, \ldots, \mathbf{H}_{k,L}]$, $\mathbf{H}_{k,l}$ channel between user $l$ and relay $k$
  - $\mathbf{N}_k \sim \mathcal{CN}(0, \mathbf{\Sigma}_k)$ is AWGN noise at relay $k$, assumed independent

- Outputs satisfy $Y_k -\!\!\circ\!\!- X_{\mathcal{L}} -\!\!\circ\!\!- Y_{\mathcal{K}\setminus k}$

- Theorem 1 characterizes its **capacity region**. Finding the optimal $U_1, \ldots, U_K$ is difficult

# Capacity under Gaussian Signaling and Enabled Resource-Sharing

**Theorem (Capacity Region under Gaussian Input with Enabled Resource-Sharing)**

*Let the input vectors use Gaussian Signaling with Enabled Resource-Sharing, i.e.,*

$$\mathbf{X}_{l,q} \sim \mathcal{CN}(0, \mathbf{K}_{l,q}) \quad q \in \{1, \dots, |\mathcal{Q}|\} \qquad \sum_{q\in\mathcal{Q}} p_Q(q)\mathbf{K}_{l,q} \preceq \mathbf{K}_l$$

*The capacity region is given by the set of all rate tuples* $(R_1, \dots, R_L)$ *satisfying that for all* $\mathcal{T} \subseteq \mathcal{L}$ *and all* $\mathcal{S} \subseteq \mathcal{K}$

$$\sum_{t\in\mathcal{T}} R_t \leq \sum_{k\in\mathcal{S}} \left[ C_k - \mathrm{E}_Q \log \frac{|\mathbf{\Sigma}_k^{-1}|}{|\mathbf{\Sigma}_k^{-1} - \mathbf{B}_{k,q}|} \right] + \mathrm{E}_Q \log \frac{|\sum_{k\in\mathcal{S}^c} \mathbf{H}_{k,\mathcal{T}}^H \mathbf{B}_{k,q} \mathbf{H}_{k,\mathcal{T}} + \mathbf{K}_{\mathcal{T},}^{-1}|}{|\mathbf{K}_{\mathcal{T},q}^{-1}|},$$

*for some* $\mathbf{0} \preceq \mathbf{B}_{k,q} \preceq \mathbf{\Sigma}_k^{-1}$*, where* $\mathbf{H}_{k,\mathcal{T}}$ *is the channel between* $\mathbf{X}_{\mathcal{T}}$ *and* $\mathbf{Y}_k$*.*

- Extends *[Theorem 5, Sanderovich et al'09]* to $L$ users and MIMO.
- Achievable with $\mathbf{U}_{k,q} = \mathbf{Y}_{k,q} + \mathbf{Z}_{k,q}$, $\mathbf{Z}_{k,q} \sim \mathcal{CN}(\mathbf{0}, \mathbf{B}_k^{-1} - \mathbf{\Sigma}_{k,q})$, $q \in \mathcal{Q}$
- Gaussian signaling can be strictly suboptimal *[Sanderovich et al'09]*

# Converse Part

- For $(X, U)$ arbitrarily correlated,

$$\log|(\pi e)\ \mathbf{J}^{-1}(X|U)| \leq h(X|U) \leq \log|(\pi e)\ \mathrm{mmse}(X|U)|$$

- For each $Q = q$,

$$I(\mathbf{Y}_k; \mathbf{U}_k|\mathbf{X}_{\mathcal{L}}, Q = q) = \log|(\pi e)\mathbf{\Sigma}_k| - h(\mathbf{Y}_s|\mathbf{X}_{\mathcal{L},q}, \mathbf{U}_{s,q}, Q = q)$$
$$\geq \log \frac{|\mathbf{\Sigma}_k^{-1}|}{|\mathbf{\Sigma}_k^{-1} - \mathbf{B}_{k,q}|},$$

where $\mathbf{0} \preceq \mathbf{B}_{k,q} \preceq \mathbf{\Sigma}_k^{-1}$ is chosen such that

$$\mathrm{mmse}(\mathbf{Y}_k|\mathbf{X}_{\mathcal{L},q}, \mathbf{U}_{k,q}) = \mathbf{\Sigma}_k - \mathbf{\Sigma}_k\mathbf{B}_{k,q}\mathbf{\Sigma}_k$$

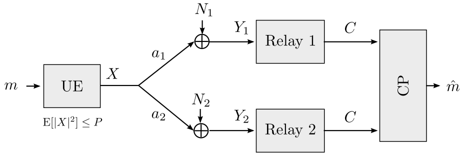
- Also,

$$I(\mathbf{X}_{\mathcal{T}}; \mathbf{U}_{S^c}|\mathbf{X}_{\mathcal{T}^c}, Q = q) = h(\mathbf{X}_{\mathcal{T},q}|\mathbf{X}_{\mathcal{T}^c}, q) - h(\mathbf{X}_{\mathcal{T}}|\mathbf{X}_{\mathcal{T}^c}, \mathbf{U}_{S^c,q}, Q = q)$$
$$\leq \log|\mathbf{K}_{\mathcal{T},q}| + \log\left|\sum_{k\in\mathcal{S}^c} \mathbf{H}_{k,\mathcal{T}}^H\mathbf{B}_{k,q}\mathbf{H}_{k,\mathcal{T}} + \mathbf{K}_{\mathcal{T},q}^{-1}\right|$$

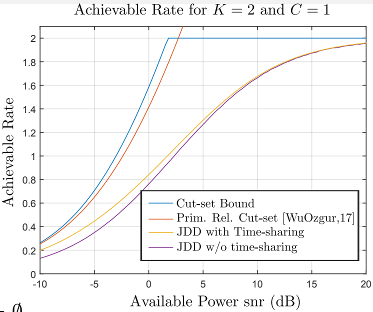
by deBrujin Identity *[Palomar-Verdu'06],[Ekrem-Ulukuss'14], [Zhou et al'17]*

$$\mathbf{J}(\mathbf{X}_{\mathcal{T},q}|\mathbf{X}_{\mathcal{T}^c,q}, \mathbf{U}_{S^c,q}) = \sum_{k\in\mathcal{S}^c} \mathbf{H}_{k,\mathcal{T}}^H\mathbf{B}_{k,q}\mathbf{H}_{k,\mathcal{T}} + \mathbf{K}_{\mathcal{T},q}^{-1}.$$

# Resource-sharing Enlarges Capacity Region




$Y_k = a_k X + N_k$, with $\mathrm{E}[X^2] \leq P$ and $N_k \sim \mathcal{N}(0,1)$, $k = 1, 2$




- JDD without resource (time)-sharing, i.e., $Q = \emptyset$

$$R_{\mathrm{JDD-w/o-ts}}(C,P) = \frac{1}{2}\log\left(1 + 2a^2Pe^{4C} + a^2P - \sqrt{(a^2P)^2 + (1+2a^2P)e^{4C}}\right)$$

- JDD with resource (time)-sharing ($|\mathcal{Q}| = 2$. Recall that $|\mathcal{Q}| \leq K + 2 = 4$ here)
  - Phase I: UE transmits at $P/\alpha$ for $\alpha n$ samples. Relays compress at $C_k/\alpha$.
  - Phase II: UE and Relays remain inactive for $(1-\alpha)n$ remaining samples.
- Intuition: For small $P$, the observations at the relays are too noisy; and, so, it is more advantageous to increase power and compression rate during shorter time.
- Improved (cut-set) upper bound: improved (cut-set) upper bound for the primitive relay channel of [Wu-Ozgur IT2018, arXiv:1701.02043v2] for our setup with $C_1 = \infty$.

# Capacity under Constant Gaussian Signaling

**Theorem (Capacity Region under Constant Gaussian Input)**

*If the input vectors use constant Gaussian Signaling, i.e.,*

$$\mathbf{K}_{1,l} = \cdots = \mathbf{K}_{|\mathcal{Q}|,l} = \mathbf{K}_l, \qquad \mathbf{X}_l \sim \mathcal{CN}(0, \mathbf{K}_l),$$

*the capacity region is given by the set of all rate tuples $(R_1, \ldots, R_L)$ satisfying that for all $\mathcal{T} \subseteq \mathcal{L}$ and all $\mathcal{S} \subseteq \mathcal{K}$*

$$\sum_{t \in \mathcal{T}} R_t \leq \sum_{k \in \mathcal{S}} \left[ C_k - \log \frac{|\mathbf{\Sigma}_k^{-1}|}{|\mathbf{\Sigma}_k^{-1} - \mathbf{B}_k|} \right] + \log \frac{|\sum_{k \in \mathcal{S}^c} \mathbf{H}_{k,\mathcal{T}}^H \mathbf{B}_k \mathbf{H}_{k,\mathcal{T}} + \mathbf{K}_{\mathcal{T}}^{-1}|}{|\mathbf{K}_{\mathcal{T}}^{-1}|},$$

*for some $\mathbf{0} \preceq \mathbf{B}_k \preceq \mathbf{\Sigma}_k^{-1}$, where $\mathbf{H}_{k,\mathcal{T}}$ is the channel between $\mathbf{X}_{\mathcal{T}}$ and $\mathbf{Y}_k$.*

- Resource-sharing at the relays does not enlarge the capacity region under constant Gaussian Signaling.
- Proof follows from Jensen's Inequality and concavity of log-det.

## Capacity under Gaussian Signaling in the High SNR Regime

- High SNR regime model:

$$\boldsymbol{\Sigma}_k = \epsilon \tilde{\boldsymbol{\Sigma}}_k; \quad \text{for some } \tilde{\boldsymbol{\Sigma}}_k \succeq \mathbf{0}, \quad \text{and } \epsilon \to 0.$$

- We have $\mathcal{R}_{\mathrm{GNS}}(C_{\mathcal{K}}) \subset \mathcal{R}_{\mathrm{GTS}}(C_{\mathcal{K}})$, where:
  - $\mathcal{R}_{\mathrm{GTS}}$: Capacity under Gaussian Input with Enabled Resource-Sharing
  - $\mathcal{R}_{\mathrm{GNS}}$: Capacity under Gaussian Input without Resource-Sharing ($Q = \emptyset$)

**Theorem (Capacity Region under Gaussian Input in High SNR)**

*If* $(R_1, \ldots, R_L) \in \mathcal{R}_{\mathrm{GTS}}(C_{\mathcal{K}})$, *then for any* $\epsilon > 0$, *for some* $\Delta_\epsilon \geq 0$,

$$(R_1 - \Delta_\epsilon, \ldots, R_L - \Delta_\epsilon) \in \mathcal{R}_{\mathrm{GNS}}(C_{\mathcal{K}})$$

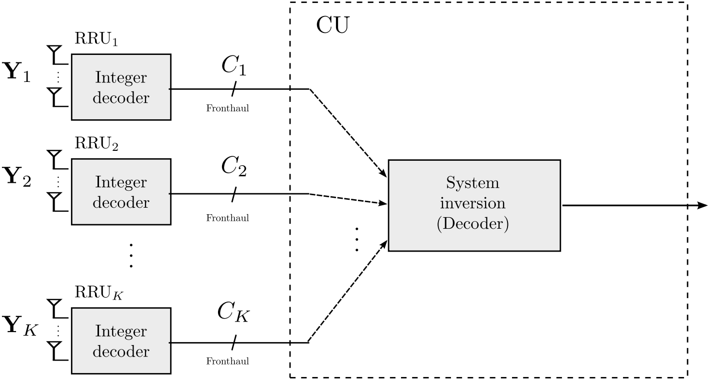
*In addition,* $\mathcal{R}_{\mathrm{GNS}}(C_{\mathcal{K}}) = \mathcal{R}_{\mathrm{GTS}}(C_{\mathcal{K}})$ *as* $\epsilon \to 0$, *since*

$$\lim_{\epsilon \to 0} \Delta_\epsilon = 0.$$

- For large SNR, the gains due to resource-sharing become limited.

# Outline

# Compute-and-Forward




- Decode linear integer equations on the transmitted messages. The equations need to be linearly independent
- At the CU, simply invert the system of equations.
- Studied in *[Nazer et al 09] [Hong and Caire 11]*.

# Compute-and-Forward (cont'd)

- Each RRU decodes an appropriate (modulo-)sum, with integer weights of the codewords transmitted by the UEs.
  - Then sends a bit stream on the FH link that identifies the decoded codeword within the lattice code.
- Upon receiving sufficient number of linear combinations, the CP can invert the resulting linear system and recover the transmitted codewords.
- For single-antenna system with $L = K$, and $C_1 = \cdots = C_K = C$, the achievable rate per UE is

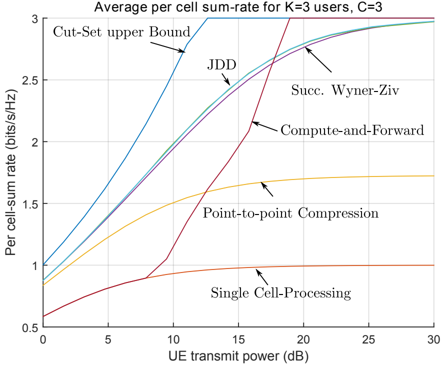
$$R_l = \min\{C, \min_{k:a_{k,l}\neq 0} R(\mathbf{h}_l, \mathbf{a}_l, \mathrm{SNR})\},$$

  where the computational rate is given by

$$R(\mathbf{h}, \mathbf{a}, \mathrm{SNR}) = \log^+\left(\frac{\mathrm{SNR}}{\mathbf{a}^H(\mathrm{SNR}^{-1}\mathbf{I} + \mathbf{h}\mathbf{h}^H)^{-1}\mathbf{a}}\right).$$

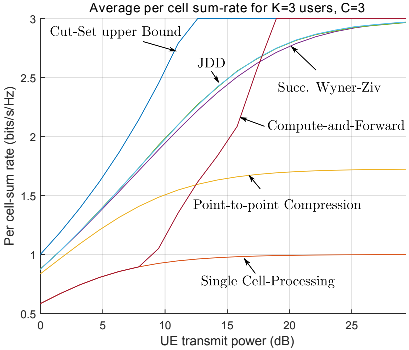
- Integer penalty: The signal received at each RRU is sum with non-integer weights of the codewords transmitted by EUs.

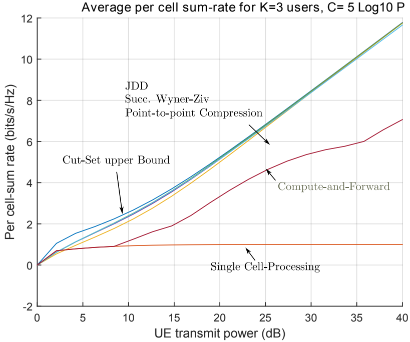
# Numerical example: Circular Wyner Model (cont'd)




- At low SNR, its performance coincides with single-cell processing.
- RUs tend to decode trivial combinations.
- At high SNR, the fronthaul capacity is the main performance bottleneck, so CoF shows the best performance.

# Cost of Obliviousness




- **Optimal degrees-of-freedom:** when fronthaul capacity grows with SNR, e.g., $C = 5\log_{10}(\mathrm{snr})$. *[Sanderhovich et al'09].*
- Capacity under Gaussian signaling to within a **constant gap** of cut-set bound.
  - If $(R_1, \ldots, R_L)$ is within the cut-set bound, then

$$((R_1 - \Delta)^+, \ldots, (R_L - \Delta)^+), \qquad \Delta \le \begin{cases} \frac{N}{2}(2.45 + \log(\frac{KM}{N})) & \text{for } KM > 2N, \\ \frac{KM+N}{2} & \text{for } KM \le 2N \end{cases}$$

# Outer Bound for General Model

**Theorem (Bounds)**

*For general DM CRAN channels with oblivious relay processing and enabled resource-sharing, a rate tuple* $(R_1, \ldots, R_L)$ *is achievable if (only if) for all* $\mathcal{T} \subseteq \mathcal{L}$ *and for all* $\mathcal{S} \subseteq \mathcal{K}$*,*

$$\sum_{t \in \mathcal{T}} R_t \leq \sum_{s \in \mathcal{S}} C_s - I(Y_S; U_S | X_{\mathcal{L}}, U_{S^c}, Q) + I(X_{\mathcal{T}}; U_{\mathcal{S}^c} | X_{\mathcal{T}^c}, Q),$$

*for some* $(Q, X_{\mathcal{L}}, Y_{\mathcal{K}}, U_{\mathcal{K}}, W)$

- *distributed according to* $P_Q \prod_{l=1}^{L} P_{X_l|Q} \; P_{Y_{\mathcal{K}}|X_{\mathcal{L}}} \; P_{W|Q}$
- $u_k = f_k(w, y_k, q)$ *for* $k = [1, K]$ *for some random variable* $W$ *and some deterministic functions* $\{f_k\}$*,* $k = [1, K]$*.*

# Remarks

- The DM CRAN problem connects with the CEO source coding problem under log-loss.
- Without $Y_k -\!\!\ominus\!\!- X_{\mathcal{L}} -\!\!\ominus\!\!- Y_{\mathcal{K}\setminus k}$ the problem is challenging even for one user:
  - Reason: Berger-Tung coding is strictly suboptimal
  - Example: Includes Korner-Marton lossless modulo-sum problem [Korner-Marton'79]
- The DM CRAN problem connects with the distributed Information Bottleneck problem studied in [Aguerri-Zaidi, IZS2018].

# Wrap Up

- We have studied transmission over a CRAN under oblivious processing constraints at the relays and enabled resource-sharing.
    - i.e., relays are not allowed to know or acquire the users' codebooks.
- Our results shed light on the optimal relay operations:
    - NNC and CoF-JD optimal when the outputs at the relay nodes are conditionally independent on the users inputs.
    - CoF-SD achieves optimal sum-rate.
    - Computed the Capacity Region under Gaussian Inputs in MIMO CRAN.
- Oblivious processing as studied relevant from a practical viewpoint:
    - Bounded rate loss in comparison with the non-oblivious setting.
    - Optimal degrees-of-freedom when fronthaul capacity grows with SNR.

# Short Outlook

- Duality issues:
  - Downlink/uplink, e.g., Compute-forward v.s. reverse Compute-forward
  - Gaussian MAC/BC duality extends also for finite-capacity fronthauls $\{C_k\}$

    - See, e.g., *Liang Liu, Pratik Patil, and Wei Yu, "An Uplink-Downlink Duality for Cloud Radio Access Network," ISIT'2016.*

- Channel/Source coding duality
  - Gelfand-Pinsker coding/Wyner-Ziv source coding:

    - *T. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information", IEEE Transactions on Information Theory, Jun. 2002.*
  - Practical implications: *Ankit Gupta, and Sergio Verdù, "Operational Duality between Gelfand-Pinsker and Wyner-Ziv Coding," ISIT 2010.*

- Optimal UE signal distribution, under average power constraint [Sanderovich et al.'09]

- Vector distributed bottleneck [Aguerri-Zaidi, IZS2018], with combination of dependent and independent components: Possible other applications.

# Some Related Tutorials


- S.-H. Park, O. Simeone, O. Sahin and S. Shamai (Shitz), “Fronthaul compression for cloud radio access networks,” IEEE Sig. Proc. Mag., Special Issue on Signal Processing for the 5G Revolution, vol. 31, no. 6, pp. 69-79, Nov. 2014.
- M. Peng, C. Wang, V. Lau and H. V. Poor, “Fronthaul-Constrained Cloud Radio Access Networks: Insight and Challenges,” IEEE Wireless Comm., vol. 22, no. 2, pp. 152-160, Apr. 2015.
- Yuhan Zhou ; Yinfei Xu ; Wei Yu ; Jun Chen, “On the Optimal Fronthaul Compression and Decoding Strategies for Uplink Cloud Radio Access Networks,”IEEE Transactions on Information Theory, vol. 62, no. 12, Dec. 2016.
- A. Zaidi and I. E. Aguerri, “Tutorial: Fronthaul Compression for Cloud Radio Access Networks,” The Thirteenth Internationl Symposium on Wireless Communication Systems (ISWCS’16), Sep. 20-23, 2016, in Pozna#, Poland.
- O. Simeone, S.-H. Park, O. Sahin and S. Shamai (Shitz), “Frontal Compression for C-RAN,” Chapter 14 in Cloud Radio Access Networks: Principles, Technologies, and Applications, T. Q. S. Quek, M. Peng, O. Simeone, and W. Yu, Eds. Cambridge University Press, Feb. 2017.
- Z. Guizani and N. Harmdi, ”CRAN, H-CRAN, and F-RAN for 5G systems: Key capabilities and recent advances,” International Journal of Network Management, pp. 1-22, 2017.

Thank you!

**Shlomo Shamai**
The Andrew and Erna Viterbi Department of Electrical Engineering, Technion-Israel Institute of Technology.

**Title:** "On the Capacity of Oblivious Cloud Relay Access Networks,"'

**Abstract:** In this overview talk we will address Cloud Radio Access Networks and put focus on the uplink setting, operating in an oblivious (nomadic) mode. Specifically, we present networks in which users send information to a remote central destinations through relay nodes (radio units) that are connected to the destination (central processor) via finite-capacity error-free links. The relays are constrained to operate without knowledge of the users' codebooks, i.e., they operate in an oblivious manner. The central processor, however, is informed about the users' codebooks, and attempts to decode the users' information. In particular, we establish a single-letter characterization of the capacity region of this model for a class of discrete memoryless channels in which the outputs at the relay nodes are independent given the users' inputs. We show that both relaying 'a-la Cover-El Gamal, i.e., compress-and-forward with joint decompression and decoding, and quantize-map-forward or noisy network coding, are optimal. The new converse part establishes, and utilizes, connections with the Chief Executive Officer (CEO) source coding problem under a logarithmic loss distortion measure. Memoryless vector Gaussian channels are also investigated and the capacity under Gaussian signaling, is established. For general memoryless models (i.e., networks in which relay outputs are arbitrarily correlated among them, and with the channel inputs), we develop inner and outer bounds on the capacity region. Comparisons with unconstrained operation of the relays will also be presented gaining insights to the penalty associated with the oblivious processing.

Joint work with Abdellatif Zaidi, Inaki Estella Aguerri and Giuseppe Caire